DOCUMENT RESUME

ED 161 933 TM 007 960

AUTHOR Plake, Barbara S.; And Others
TITLE An Investigation of Differential Item Performance by Sex on the Iowa Tests of Basic Skills.
PUB DATE 28 Mar 78
NOTE 37p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada, March 28, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Achievement Tests; *Content Analysis; Elementary Education; *Elementary School Mathematics; *Item Analysis; *Sex Differences; Sex Discrimination; *Test Bias; Test Items
IDENTIFIERS Iowa Tests of Basic Skills

ABSTRACT

Three levels of the Iowa Tests of Basic Skills were studied to disclose the possible existence of sex bias in mathematics test items. Two mathematics tests (mathematical concepts and mathematics problem solving) and two comparison verbal tests (vocabulary and reading) were selected for analysis at three levels--grades 3, 6, and 8. Samples of 480 boys and 480 girls were selected from 48 Iowa school systems at each grade level. A three-factor analysis of variance design was used with items and sex considered fixed factors, and school systems a random factor. Out of 341 verbal items, there were 13 that appeared to favor girls and there were 16 that appeared to favor boys. Out of 213 mathematics items, there were 8 that appeared to favor girls and 11 that appeared to favor boys. The results of the study suggest that neither the verbal nor the mathematical subtests function differentially for male and female pupils at any of the three grade levels tested. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED161933

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Barbara S Plake
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

An Investigation of Differential Item Performance by Sex on the Iowa Tests of Basic Skills

Barbara S. Plake
University of Nebraska

H. D. Hoover and Brenda H. Loyd
University of Iowa

Paper presented at
Annual Meeting of
the National Council on Measurement in Education
Toronto, Ontario
March 28, 1978

TM007 960

An Investigation of Differential Item Performance by Sex on the Iowa Tests of Basic Skills

Sex bias has become an increasingly important consideration in standardized educational tests. Especially in the area of mathematics, there has been much research and debate on how sex differences are expressed in differences in achievement in mathematics. Sells (1976) has demonstrated that mathematical achievement in high school acts as a critical factor affecting university admissions and as a filter restricting the entrance of proportionate numbers of some subgroups into professional areas with specific mathematical prerequisites. Therefore appropriately assessing mathematical achievement has important social consequences.

At the elementary school level, much of the research (Mowder and Chase, 1977, King and Blount, 1975) into sex bias in mathematics has focused on teacher developed tests. These tests were constructed to emphasize stereotypical male or female activities in the presentation of mathematics problems in an attempt to assess potential sex bias which may result from the language or context of mathematics problems. The findings from such studies have been neither consistent nor conclusive. There remains the need to investigate sex differences in performance on mathematics problems presented with a minimum of language as well as in "worded" mathematics problems. In addition, there exists the need for more thorough investigation of differential performance by sex on nationally standardized achievement tests. The selection of nationally standardized achievement tests for study avoids many problems associated with teacher developed tests, and because of the wide use and influence

of such tests, the assessment of sex differences in performance on such tests becomes even more necessary and important.

One of many researchers exploring sex differences in mathematics achievement, Dwyer (1976) considered sex in relation to test content in mathematics She found that many researchers have reported their findings in a way which tends to support the traditional view that boys are superior in mathematics and girls in verbal skills, but when these findings are examined more closely, they are often inconclusive in showing a clear difference between mathematical achievement for boys and girls. Furthermore, sex role stereotyping and male/female representation in test items have often been ignored in mathematics tests. It is possible that a small number of individual test items could modify the pattern of sex differences reported.

The Manual for Administrators, Supervisors, and Counselors for the Iowa Tests of Basic Skills (ITBS) (Hieronymus et al., 1974) reports sex differences in achievement on the ITBS at three levels of achievement for a representative national sample. Differences at three points in the distribution are examined: the 10th percentile, the 50th percentile and the 90th percentile. For all three achievement points, the difference in verbal skills is large and favors girls. These differences tend to increase with grade level. In mathematics the middle and lower percentile groups show a slight advantage for girls which remains constant across grade level (from third through eighth grade). For upper percentile groups, the differences reverse to favor boys and increase from grade three through grade eight.

Hilton and Berglund (1974) conducted a longitudinal study to assess sex differences in mathematics achievement as measured by STEP

Math and SCAT-Q. Their results indicate no significant sex differences in achievement at the fifth grade level, but at subsequent grade levels (7, 9, 11) males had higher scores than females and the difference increased with increase in age.

Fennema (1974) reviewed educational literature pertaining to mathematics achievement and sex differences and concluded that even across divergent studies there are no significant differences that "consistently" appear between boys and girls in the fourth to ninth grade. When differences do appear they support the contention that girls perform better on tests of mathematical computation and boys tend to perform better on tests of mathematical reasoning.

To the extent that sex differences in achievement are perceived in the results of standardized tests, the tests themselves must also become the objects of investigation. Dwyer (1976) suggests two major types of test content that may influence sex differences on standardized tests. The first is test content related to sex-role stereotyping and sexism; that is, the proportional representation of males and females and the roles they occupy throughout the items of a test. The second is test content related to psychometric considerations: item types (items which measure different aspects of mathematics curriculum), cognitive skills required (computation or concept analysis), and speededness or lack of it in a testing situation.

Steckler, McCarthy, and Tittle (1974) quantified sex bias in standardized tests by counting the number of masculine nouns and pronouns and the number of feminine nouns and pronouns which appear in eight standardized tests and determined that male references far outnumbered female references. Milton (1959) found that when characteristics of

math problems were altered to make the content less appropriate to masculine roles, the sex differences in problem solving was reduced.

Dwyer (1976) concluded that because a small difference in raw score means may be statistically significant, one or two biased items could lead to such a difference, and that the test content factors should therefore be examined at the item level.

Coffman (1961) analyzed sex differences in the item responses of men and women who took the Scholastic Aptitude Test (SAT) in 1954. He investigated the question of equal item difficulties for groups of men versus women. An item was considered possibly biased if item difficulties for the two groups showed a difference that was significantly greater than would be expected by chance. An arcsin transformation was used to standardize the variance of item difficulties. Prior to completion of the item analysis, experts examined the items and made predictions about which items they felt would favor men and which would favor women. The analysis showed the predictions to be statistically significant at the .05 level, with fourteen of the seventeen predictions substantiated; six of the items were easier for men, and eight were easier for women.

Three studies have contributed significantly to the development of methodology used in this study.

Cardall and Coffman (1964) expanded the research on item bias in their investigation of the appropriateness of the SAT for three subgroups. They used an analysis of variance technique and an arcsin transformation to standardize the variance of the item difficulties. They considered a significant item by group interaction to be an indication of possible bias. Their method has the advantage of separating the overall differences of the subgroups from the differences attributable to individual items.

Item bias on the Preliminary Scholastic Aptitude Test for different racial and socio-economic (SES) groups was the subject of an investigation by Cleary and Hilton (1968). They considered an item "biased" for members of a particular group if the members of that group obtained an average score on that item which differed from the average score of other groups by more or less than would be expected based on their performance on other items on the same test. They employed a three-factor analysis of variance method with the SES factor nested within race. Bias was defined as an interaction between items and race, or between items and SES. Although both interactions were statistically significant, they contributed minimal percentages to the total variance of the observed scores. Follow-up analysis led to the conclusion that the test was not biased for the groups studied.

Plake (1976) investigated differences in item difficulty on the ITBS for pupils tested out-of-level and in-level, by using a three-factor analysis of variance technique. Group (in- or out-of-level), items, and school systems were the three independent variables. When a significant item by group interaction was obtained, a follow-up procedure was used which examined the interaction contrast for each individual item to identify specific items which were contributing to the interaction.

Procedure:

The Iowa Tests of Basic Skills (ITBS), a general achievement battery, consists of eleven subtests. Four subtests were selected for study: Vocabulary, Reading, Mathematics Concepts, and Mathematics Problem Solving. Primary attention was focused on the two mathematics subtests in order to explore the question of sex differences on mathematics tests. The Vocabulary and Reading subtests were selected to provide a basis for comparison and contrast.

The ITBS multilevel battery measures achievement of students from early elementary school through junior high school. The tests consist of six overlapping Levels, 9 through 14, corresponding roughly to grades 3 through 8, respectively. In order to explore the prospect of a developmental pattern of sex differences three levels were selected for analysis: Level 9, Level 12, and Level 14.

The data for this study were obtained from 48 school systems in Iowa, randomly selected from the Iowa school systems which tested their third, sixth, and eighth grade students with the ITBS during the 1975-76 school year. From each of the 48 school systems 10 girls and 10 boys were randomly chosen in each of the three selected grades (levels). Item responses for each student were recorded on each of the four subtests by using a "1" to indicate a correct response, and "0" to represent an incorrect response or an omit.

A three-factor analysis of variance design was used to analyze the results. The first factor, items, and the second factor, sex, were both considered fixed factors. The third factor, school systems, was considered a random factor.

The linear model for an individual score in this design is:

$$E(X_{ijkn}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \delta_{jkn} + (\alpha\delta)_{ijkn} + \xi_{ijkn}$$

where:

X_{ijkn} = the score on the i^{th} item in the j^{th} (sex) group and the k^{th} system for the n^{th} subject

μ = the grand mean

α_i = the item effect

β_j = the sex effect

γ_k = the system effect

$(\alpha\beta)_{ij}$ = the interaction of the i^{th} item with the j^{th} sex group

δ_{jkn} = the subject effect

$(\alpha\delta)_{ijkn}$ = the interaction of subject condition

$(\alpha\gamma)_{ik}$ = the interaction of the i^{th} item with the k^{th} system

$(\beta\gamma)_{jk}$ = the interaction of the j^{th} sex group with the k^{th} system

$(\alpha\beta\gamma)_{ijk}$ = the interaction of the i^{th} item with the j^{th} sex group and the k^{th} system

ξ_{ijk} = the error

The primary hypothesis of interest was:

H_o: $\theta^2_{\alpha\beta} = 0$. There is no interaction between the items and sex groups, where $\theta^2_{\alpha\beta}$ is the items by sex effect.

An analysis of variance procedure was used to test this hypothesis for each of the four subtests at each of the three levels.[1] When a significant interaction was obtained, the interaction cortrast for each item of the subtest was computed (Plake and Hoover, 1977). The contrast for each item was of the following form:

$$\Psi_{ij} = p_{ij} - \bar{p}_i - \bar{p}_j + \bar{\bar{p}}$$

where

p_{ij} = the average item difficulty of the i^{th} item for the j^{th} sex group across the K systems

$\bar{p}_i$ = the mean of the item difficulties of the i^{th} item across the two groups and the K systems

$\bar{p}_j$ = the mean of the item difficulties for the j^{th} sex group across all items and the K systems

$\bar{\bar{p}}$ = the overall item mean

Each contrast was tested for an α_{pc} level, $\alpha_{pc} = 1 - (1 - \alpha)^{1/q}$, where α was the experimentwise level (.05) and q was the number of contrasts

to be tested (the number of items in the specific subtest). A significant contrast identified a particular item as one which contributed significantly to the items by sex interaction.

An item so designated was examined in an attempt to ascertain a plausible reason for the item's interaction characteristic. The number and type of items identified were compared across grade levels for each of the four subtests in order to discern any developmental trends. The items were also compared across subtests. Finally the percent of total variance accounted for by the item by sex interaction was calculated.

Results:

The summary tables for the analyses of variance are presented in Appendix A.[2] As would be expected with the large sample size used in this study, even small effects were detected to be statistically significant. The item by sex interactions were statistically significant ($p < .01$) for each of the four subtests at each of the three grade levels.

Interaction contrasts for each item of each subtest were computed. For purposes of clarity and aid in descriptive understanding, the contrasts were calculated using the item difficulties rather than transformed item difficulties. The items which contributed significantly to the item by group interaction are presented in Appendix B. Included with each item are the skills classification, the sex group favored by the item, the item difficulty (p-value) for each sex group, and the value of the interaction contrast.

A brief summary by sex, subtest, and level of the number of items which contributed significantly to the item by group interactions is presented in Table 1. No consistent differences in the number of items identified across subtests, grade levels, or sex are evident.

Table 1

Summary of Items Significantly Contributing

to the Item by Sex Interaction

	Level (Grade)		
	9 (3rd)	12 (6th)	14 (8th)
VOCABULARY			
# With Significant Contrasts	5	6	4
Total # in Subtest	31	46	48
# Favoring Boys	3	3	1
# Favoring Girls	2	3	3
READING			
# With Significant Contrasts	2	4	8
Total # in Subtest	60	76	80
# Favoring Boys	2	2	5
# Favoring Girls	0	2	3
MATHEMATICS CONCEPTS			
# With Significant Contrasts	2	4	4
Total # in Subtest	30	45	48
# Favoring Boys	2	4	0
# Favoring Girls	0	0	4
MATHEMATICS PROBLEM SOLVING			
# With Significant Contrasts	0	5	4
Total # in Subtest	25	31	34
# Favoring Boys	0	3	2
# Favoring Girls	0	2	2

The items significantly contributing to the items by sex interaction in the Mathematics Concepts and Mathematics Problem Solving subtests were further analyzed with respect to stereotypical male or female context and with respect to the number of male or female nouns or pronouns present. None of the identified items in the Mathematics Concepts subtests contained references to any male or female pronouns, nouns or stereotypical situations. Six of the items in the Mathematics Problem Solving subtests were presented in a context that might be said to represent stereotypical male activities (putting up a basketball goal, concern with cars and garages, etc.). Of these six items, half favored females and half males. None of the identified items presented stereotypical female activities. A count of the number of male and female nouns and pronouns in the identified items revealed four items with male references and one with female. Of the items containing male nouns and pronouns, half favored males and half females.

To analyze item differences by sex in the mathematics subtests as a function of the area of mathematics sampled or cognitive skill required, the skills classifications of the items contributing to the interaction were considered. (The skills classification for each item is listed in the ITBS teacher's manual.) Of the nine items identified in the Mathematics Problem Solving subtest, only three showed any consistency by skills objectives in favoring one specific sex group; all three items (#42, #73, #81) were classified under the skill of multiplying and dividing fractions, and favored males. Eight other items included in the analysis under the same skills classification displayed no significant differences.

In the Mathematics Concepts subtest, ten items were identified. The only consistencies among the ten items in terms of skills classifications

were: two items favoring males, under the classification of relative values of decimals (#82, #96); and two items (#92, #123) favoring females, under the classification of equations, inequalities and number sentences. The Mathematics Concepts Tests analyzed included a total of seven items under the major classification of decimals and ten under the classification of equations, inequalities and number sentences.

While the main effect of sex was not of primary interest in this investigation, the average item difficulties of boys and girls for each subtest at each level studied are presented in Table 2.

Estimates of the variance components, and the percentage contribution of each effect to the total variance of a randomly chosen single observation (item score), are presented in Table 3. The percentage of variance accounted for by the item by sex interaction is minimal for all subtests (ranging from .13% to .84%); this indicates that the item by group interaction effect lacks practical importance.

Discussion:

The results of this study suggest that the Vocabulary, Reading, Mathematics Concepts, and Mathematics Problem Solving subtests of the ITBS do not function differentially for male and female pupils at the three grade levels tested (third, sixth, and eighth grades). The item by sex interactions were found to be statistically but not practically significant. The percentage of the total variance accounted for by the interaction components was minimal.

The results provided little support for the contention that test content as it relates to sex role stereotyping, or the frequency of male or female nouns or pronouns, affects performance on mathematics test items.

Table 2

Average Item Difficulties By Sex

	Level					
	9 (Grade 3)		12 (Grade 6)		14 (Grade 8)	
Subtest	M	F	M	F	M	F
Reading	.53	.58	.51	.52	.56	.60
Vocabulary	.60	.62	.60	.60	.56	.57
Mathematics Concepts	.65	.67	.55	.55	.53	.56
Mathematics Problem Solving	.55	.59	.51	.50	.49	.50

Table 3

Estimated Components of Variance

Vocabulary

Effect	Gr. 3 Est.	Gr. 3 % of Total	Gr. 6 Est.	Gr. 6 % of Total	Gr. 8 Est.	Gr. 8 % of Total
Groups	.00033	.14	0	0	0	0
Systems	.00330	1.38	.00169	.56	.00135	.51
Groups by Systems	.00012	.50	0	0	.00063	.24
Subject	.03034	12.64	.02454	8.13	.02543	9.65
Items	.03467	14.44	.03184	10.55	.03396	12.88
Items by Groups	.00096	.40	.00126	.42	.00084	.32
Items by Systems	.00275	1.15	.00268	.89	.00424	1.61
Items by Groups by Systems	.00027	.11	.00035	.12	.00036	.14
Residual	.16733	69.70	.23954	79.35	.19680	74.66

Reading

Effect	Gr. 3 Est.	Gr. 3 % of Total	Gr. 6 Est.	Gr. 6 % of Total	Gr. 8 Est.	Gr. 8 % of Total
Groups	.00129	.52	0	0	.00075	.31
Systems	.00376	1.51	.00140	.56	.00154	.63
Groups by Systems	0	0	0	0	.00039	.16
Subject	.03190	12.80	.02596	10.36	.02537	10.32
Items	.02958	11.88	.02176	8.68	.01743	7.10
Items by Groups	.00045	.18	.00112	.45	.00205	.84
Items by Systems	.00143	.57	.00174	.70	.00132	.54
Items by Groups by Systems	.00021	.08	.00041	.17	.00025	.10
Residual	.18047	72.45	.19831	79.10	.19680	80.03

Mathematics Concepts

Effect	Gr. 3 Est.	Gr. 3 % of Total	Gr. 6 Est.	Gr. 6 % of Total	Gr. 8 Est.	Gr. 8 % of Total
Groups	.00016	.07	0	0	.00019	.08
Systems	.00442	1.94	.00389	1.56	.00213	.86
Groups by Systems	0	0	0	0	.00052	.20
Subject	.02600	11.42	.02834	1.36	.02397	2.40
Items	.03062	13.45	.02605	10.45	.02553	10.26
Items by Groups	.00089	.391	.00087	.35	.00088	.35
Items by Systems	.00332	1.46	.00676	2.71	.00558	2.24
Items by Groups by Systems	0	0	0	0	.00200	.08
Residual	.16830	68.20	.19687	78.27	.19803	79.03

Table 3 (cont.)

Estimated Components of Variance

Mathematics Problem Solving

Effect	Gr. 3 Est.	Gr. 3 % of Total	Gr. 6 Est.	Gr. 6 % of Total	Gr. 8 Est.	Gr. 8 % of Total
Groups	.00061	.25	0	0	0	0
Systems	.00572	2.32	.00215	.85	.00097	.39
Groups by Systems	0	0	.00017	.07	.00064	.25
Subject	.03149	12.76	.03058	12.16	.02057	8.21
Items	.03736	15.14	.01898	7.55	.02777	11.08
Items by Groups	.00033	.13	.00145	.58	.00147	.59
Items by Systems	.00236	.96	.00133	.53	.00111	.44
Items by Groups by Systems	.00060	.24	0	0	.00101	.41
Residual	.16830	68.20	.19687	78.27	.19803	79.03

In the examination of the skills classification of the mathematics items contributing to the sex by items interaction, there was an indication that skills involving the relative values of decimals and multiplying and division of fractions favored boys, while skills involving equations and number sentences favored girls. However, other items in the subtests classified under the same skills objectives failed to display any differential performance by sex.

The number and type of items exhibiting sex differences on the mathematics subtests tended to be distributed evenly between both sex groups, indicating that some of these differences may be due to chance or less than obvious explanations.

FOOTNOTES

[1]The analysis of variance was also conducted employing the transformed item difficulties for each sex group from each school system. The transformation was of the form $\phi = 2 \arcsin \sqrt{p}$.

[2]Due to the balance of the design, the results of the analysis of variance using individual scores and that using transformed item difficulties were strictly parallel. Only individual score analyses are presented in this paper.

REFERENCES

Cardall, Carolyn and Coffman, W.E. A Method for Comparing the Performance of Different Groups on the Items in a Test. Research Bulletin 64-67, Princeton, N.J.: Educational Testing Service, 1964.

Cleary, T.A. and Hilton, T.L. An Investigation of Item Bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Coffman, W.E. Sex Differences in Responses to Items in an Aptitude Test. 18th Yearbook, NCME, 1961.

Dwyer, Carol Anne. Test Content in Mathematics and Science: The Consideration of Sex. AERA, 1976.

Fennema, E. Mathematics Learning and the Sexes: A Review. Journal of Educational Research in Mathematics Education, 1974, 5, 126-139.

Hilton, T.L. and Berglund, G.W. Sex Differences in Mathematics Achievement--A Longitudinal Study. Journal of Educational Research, 1974, 67, 231-237.

King, P.E. and Blount, H.P., The influence of sex-typed verbal math test problems on performance of sixth grade students. Paper presented at the annual meeting of The National Council on Measurement in Education, Washington, D.C., 1975.

Milton, G.A. Sex Differences as a Function of Role Appropriateness of Problem Content. Psychological Reports, 1959, 5, 705-708.

Mowder, B.A. and Chase, C.I., Sex biased mathematics items and performance of elementary school students. Paper presented at the annual meeting of The National Council of Measurement in Education, New York, N.Y., 1971.

Plake, B.S. and Hoover, H.D. An Analytical Method of Identifying Biased Test Items. AERA, 1977.

Plake, B.S. The Comparability of Equal Raw Scores for Children at Different Grade Levels: An Issue in "Out-of-level" Testing. Doctoral Dissertation, University of Iowa, 1976.

Sells, Lucy W. The Mathematics Filter and the Education of Women and Minorities. American Association for the Advancement of Science, Feb., 1976.

Steckler, J.F., McCarthy, K.A. and Tittle, C.K. A Quantitative Method for Measuring Sex "Bias" in Standardized Tests. Journal of Educational Measurement, 1974, 11, 151-161.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Company, 1962.

APPENDIX A

Appendix A

Analysis of Variance Summary Table (p-values)

Level 9 (Grade 3) Vocabulary

Source	df	MS	F
Between	959	1.217	
Groups	1	6.069	5.123
Systems	47	3.152	2.845
Groups by Systems	47	1.184	1.069
Error (b)	864	1.108	
Within	28800	.205	
Items	30	33.500	151.937
Items by Groups	30	.632	3.658**
Items by Systems	1410	.220	1.318
Items by Groups by Systems	1410	.175	1.033
Error (w)	25920	.167	
Total	29759	.238	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 9 (Grade 3) Reading

Source	df	MS	F
Between	959	2.345	
Groups	1	36.804	18.689
Systems	47	6.601	3.152
Groups by Systems	47	1.969	.940
Error (b)	864	2.095	
Within	56640	.212	
Items	59	28.605	136.501
Items by Groups	59	.399	2.165**
Items by Systems	2773	.209	1.161
Items by Groups by Systems	2773	.185	1.023
Error (w)	50976	.180	
Total	57599	.24746	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 9 (Grade 3) Mathematics Concepts

Source	df	MS	F
Between	959	1.065	
Groups	1	3.104	4.164
Systems	47	3.596	3.816
Groups by Systems	47	.745	.791
Error (b)	864	.942	
Within	27840	.196	
Items	29	29.629	128.257
Items by Groups	29	.579	3.792**
Items by Systems	1363	.231	1.420
Items by Groups by Systems	21363	.152	.94
Error (w)	25056	.162	
Total	28799	.225	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 9 (Grade 3) Mathematics Problem Solving

Source	df	MS	F
Between	959	1.103	
Groups	1	8.251	8.702
Systems	47	3.816	3.99
Groups by Systems	47	.948	.992
Error (b)	864	.955	
Within	23040	.209	
Items	24	36.083	165.881
Items by Groups	1124	.338	1.871**
Items by Systems	1128	.217	1.292
Items by Groups by Systems	1128	.181	1.075
Error (w)	20736	.168	
Total	23999	.245	

**p < .01

Appendix A (cont.)

Analysis of Variance Summary Table (p-values)

Level 12 (Grade 6) Vocabulary

Source	df	MS	F
Between	959	1.434	
Groups	1	.127	.109
Systems	47	2.925	2.138
Groups by Systems	47	1.170	.855
Error (b)	864	1.368	
Within	43200	.213	
Items	45	30.801	132.661
Items by Groups	45	.789	4.274**
Items by Systems	2115	.232	1.308
Items by Groups by Systems	2115	.184	1.040
Error (w)	2115	.177	
Total	44159	.239	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 12 (Grade 6) Reading

Source	df	MS	F
Between	959	2.263	
Groups	1	1.456	.716
Systems	47	4.198	1.933
Groups by Systems	47	2.035	.937
Error (b)	864	2.171	
Within	72000	.223	
Items	75	21.119	90.380
Items by Groups	75	.743	3.597**
Items by Systems	3525	.234	1.178
Items by Groups by Systems	3525	.207	1.04
Error (w)	64800	.198	
Total	72959	.249	

**p < .01

Appendix A (cont.)
Analysis of Variance Summary Table (p-values)

Level 12 (Grade 6) Mathematics Concepts

Source	df	MS	F
Between	959	1.629	
Groups	1	.000	
Systems	47	4.961	3.400
Groups by Systems	47	1.459	1.000
Error (b)	864	1.459	
Within	44240	.216	
Items	44	25.185	78.270
Items by Groups	44	.591	3.416**
Items by Systems	2068	.321	1.754
Items by Groups by Systems	2068	.173	.943
Error (w)	38016	.83	
Total	43199	.247	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 12 (Grade 6) Mathematics Problem Solving

Source	df	MS	F
Between	959	1.214	
Groups	1	.141	.114
Systems	47	2.476	2.163
Groups by Systems	47	1.248	1.090
Error (b)	864	1.144	
Within	28800	.218	
Items	30	18.445	82.219
Items by Groups	30	.891	4.570**
Items by Systems	1410	.224	1.139
Items by Groups by Systems	1410	.195	.990
Error (w)	25920	.197	
Total	29759	.249	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 14 (Grade 8) Vocabulary

Source	df	MS	F
Between	959	1.495	
Groups	1	1.876	.936
Systems	47	2.696	1.924
Groups by Systems	47	2.004	1.430
Error (b)	864	1.401	
Within	45120	.219	
Items	47	32.872	123.048
Items by Groups	47	.593	3.158**
Items by Systems	2209	.267	1.48
Items by Groups by Systems	2209	.188	1.04
Error (w)	40608	.180	
Total	46079	.246	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 14 (Grade 8) Reading

Source	df	MS	F
Between	959	2.409	
Groups	1	32.054	11.224
Systems	47	4.689	2.106
Groups by Systems	47	2.855	1.282
Error (b)	864	2.226	
Within	75840	.216	
Items	79	16.956	75.853
Items by Groups	79	1.187	5.885**
Items by Systems	3713	1.223	1.136
Items by Groups by Systems	3713	.202	1.025
Error (w)	68256	.197	
Total	76799	.244	

**p < .01

Level 14 (Grade 8) Mathematics Concepts

Source	df	MS	F
Between	959	1.469	
Groups	1	6.211	3.419
Systems	47	3.388	2.527
Groups by Systems	47	1.816	1.355
Error (b)	864	1.341	
Within	45120	.222	
Items	47	24.821	81.623
Items by Groups	47	.617	3.162**
Items by Systems	2209	.304	1.599
Items by Groups by Systems	2209	.195	1.027
Error (w)	40608	.190	
Total	46079	.248	

**p < .01

Analysis of Variance Summary Table (p-values)

Level 14 (Grade 8) Mathematics Problem Solving

Source	df	MS	F
Between	959	.951	
Groups	1	.566	.426
Systems	47	1.560	1.738
Groups by Systems	47	1.329	1.481
Error (b)	864	.897	
Within	31680	.229	
Items	33	26.878	121.637
Items by Groups	33	.926	4.226**
Items by Systems	1551	.221	1.116
Items by Groups by Systems	1551	.219	1.106
Error (w)	28512	.198	
Total	32639	.250	

**p < .01

APPENDIX B

ITBS ITEMS
with Significant Sex Differences

Vocabulary

	Level 9 (Grade 3) [31 items]			Group Favored (P_M, P_F, Ψ)
12.	Approach the zoo			
	1) come near	3)	walk into	M
	2) leave behind	4)	try to find	(.525, .469, .042)
13.	Puddle of water			
	1) pail	3)	stream	M
	2) spot	4)	small pool	(.679, .621, .043)
21.	Cross the plains			
	1) hills	3)	level land	M
	2) rivers	4)	deep valleys	(.481, .373, .068)
22.	Expecting an answer			
	1) waiting for	3)	offering	F
	2) listening to	4)	giving	(.606, .721, .043)
27.	A cozy room			
	1) comfortable	3)	tiny	F
	2) neat	4)	drafty	(.637, .775, .054)

	Level 12 (Grade 6) [46 items]			
46.	Knit a scarf			
	1) make with needles	3)	try to buy	F
	2) sew together	4)	line with fur	(.729, .871, .069)
55.	Bad behavior			
	1) reputation	3)	relations	M
	2) thoughts	4)	conduct	(.694, .560, .068)
59.	Fragrance of grapes			
	1) taste	3)	preparation	F
	2) appearance	4)	odor	(.812, .910, .047)

62. An antique watch
1) unusual 3) old
2) accurate 4) expensive
F (.723, .837, .055)

71. Maneuver into line
1) step backward 3) march
2) move skillfully 4) blunder
M (.708, .600, .056)

73. Zenith of power
1) end 3) peak
2) evil 4) limitation
M (.358, .256, .053)

Level 14 (Grade 8) [48 items]

73. Zenith of power
1) end 3) peak
2) evil 4) limitation
M (.660, .550, .062)

79. A grave expression
1) solemn 3) strange
2) fatal 4) creative
F (.506, .666, .073)

104. Recite in unison
1) together 3) one at a time
2) loudly 4) in turn
F (.625, .745, .054)

106. Barely legible
1) legal 3) understandable
2) readable 4) possible
F (.406, .510, .045)

ITBS ITEMS
with Significant Sex Differences

Reading Comprehension
(Reading passages not included)

	Skills Classification/ Group Favored (P_M, P_F, Ψ)
Level 9 (Grade 3) [60 items]	
11. What is a fishing shack? 1) A building 2) A fish pond 3) A hole in the ice 4) A line on which to catch fish	Recognize and understand important facts and details/M (.743, .692, .051)
20. What does Mr. Kozek do when he "trims" trees? 1) Decorates trees 2) Waters trees 3) Cuts off tree branches 4) Cuts down trees	Deduce the meaning of words or phrases from context/M (.727, .654, .061)
Level 12 (Grade 6) [76 items]	
66. A "tributary" is which of these? 1) A river that runs into the ocean 2) A stream that runs into another stream 3) A river that runs a great distance 4) An ocean that is fed by a river	Deduce the meaning of words or phrases from context/M (.664, .550, .061)
80. Why was this area called the "Unassigned Lands"? 1) It was not being put to any legal use. 2) Indian tribes refused assignment there. 3) It was useless land. 4) The government had overlooked the land.	Deduce the meaning of words or phrases from context/F (.556, .675, .055)
111. A full-grown blue whale would be most like which of these in length? 1) A fishing rod 2) A large house 3) A rowboat 4) Three city blocks	Recognize and understand implied facts and relationships/M (.406, .262, .076)

135. How does the poet seem to feel about the first travelers of this road?
1) Sympathetic and admiring of their courage
2) Surprised at their daring
3) Angry at their foolishness
4) Disgusted with their ignorance

Recognize the writer's viewpoint, attitude or intention/F (.470, .591, .056)

Level 14 (Grade 8) [80 items]

102. In what way is trash a danger to health?
1) It provides homes for rats.
2) It drives away visitors.
3) It harms people's lungs.
4) It is an eyesore.

Recognize and understand implied facts and relationships/M (.527, .435, .066)

118. Which would be the best headline for this article if used as a newspaper story?
1) "The Great Blue Whale Is Extinct"
2) "International Whaling Commission Organized"
3) "The Great Blue Whale in Peril"
4) "Products of the Whaling Industry"

Recognize the main idea or topic of a paragraph or selection/M (.608, .548, .050)

120. Which of these might be partly responsible for the improvement in whaling techniques referred to in the fourth sentence?
1) The invention of the whaling harpoon
2) The development of radar
3) The formation of the International Whaling Commission
4) New laws that allow the use of large guns in whaling

Develop generalizations from a selection/M (.356, .266, .065)

124. Which of these was a result of building the Grand Coulee Dam?
1) Several large cities were moved or razed.
2) Many miles of streets and roads went under water.
3) A great number of railroad trains were destroyed.
4) Many livestock were drowned.

Detect main purpose of a paragraph or selection/M (.654, .477, .109)

143. What was the purpose of the thermometer set in the milk?
1) To stir it
2) To show when it was cool enough
3) To show that it was warm enough to dissolve the rennet
4) To show how much coloring was needed

Recognize and understand implied facts and relationships/M (.385, .327, .049)

173. In teaching our alphabet, we sometimes say "A is for Apple." How might the Phoenicians have used this principle?
1) is for bull
2) is for ox
3) is for Apis
4) is for Aleph

Develop generalization from a selection/F (.477, .623, .052)

174. Which of these topics is discussed in the second paragraph?
1) The development of the letter A
2) How the Romans changed the alphabet
3) The origin of the second letter of the alphabet
4) Why the Greeks reversed Phoenician letters

Recognize the main idea or topic of a paragraph or selection/F (.537, .696, .058)

178. Which of the following best describes the background of the word alphabet?
1) It came to us directly from the Egyptian.
2) It is made up of two Phoenician words.
3) It is the Roman word for the first two letters.
4) It is the combination of the Greek words for A and B.

Recognize and understand important facts and details/F (.479, .631, .055)

ITBS ITEMS
with Significant Sex Differences

Mathematics Concepts

Level 9 (Grade 3) [30 items]

Skills Classification/ Group Favored ($\bar{P}_M$, $\bar{P}_F$, Ψ)

12. How would you write 20 + 100 + 6 as a 3-digit numeral?
1) 800 3) 216
2) 306 4) 126

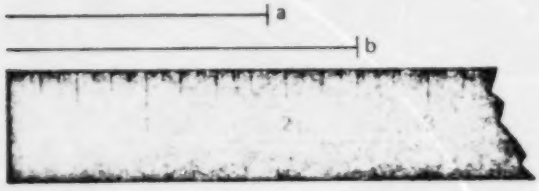
Numeration and Number Systems--place value and expanded notation/M (.614, .485, .075)

26. How many sixes must be subtracted from 24 to get 0?
1) 3 3) 5
2) 4 4) 6

Properties of numeration and number systems/M (.537, .466, .046)

Level 12 (Grade 6) [45 items]

62. In the picture below how many inches longer is line segment b than line segment a?



1) $2\frac{1}{2}$ 3) $\frac{1}{2}$
2) $1\frac{7}{8}$ 4) $\frac{5}{8}$

Measurement-Length/M (.562, .435, .062)

64. Which of these fractional numbers is greater than one-half?
1) $\frac{1}{3}$ 3) $\frac{1}{4}$
2) $\frac{2}{3}$ 4) $\frac{2}{5}$

Fractions-Relative values/M (.618, .506, .054)

82. Which of these numerals represents the greatest number?
1) .307 3) 3.07
2) .730 4) 3.7

Decimals-Relative values/M (.439, .342, .047)

96. Which of the following is greater in value than .761?
1) .8 3) .7603
2) .699 4) .716

Decimals-Relative values/M (.306, .197, .052)

Level 14 (Grade 8) [48 items]

92. What numeral should replace n in the number sentence $\frac{3}{4} \times n = \frac{2}{3}$?

1) $\frac{1}{2}$ 2) $\frac{5}{7}$ 3) $\frac{8}{9}$ 4) $1\frac{5}{12}$

Equations, Inequalities, and Number Sentences/F (.462, .575, .052)

103. In the exercise 3046 x 2.7 = 82.242, the decimal point is omitted in the name for the first factor. What is the correct factor?
1) .3046 3) 30.46
2) 3.046 4) 304.6

Decimals--Fundamental operations: ways to perform/F (.604, .729, .059)

114. What should replace n in the equation $(4 \times 5) + (4 \times 7) = n \times (5 + 7)$?

1) 4 2) 5 3) 7 4) 8

Properties of numeration and number sentences/F (.706, .844, .065)

123. If a + 16 = b, what is b - a?
1) 16 3) b - 16
2) a - 16 4) 16 - a

Equations, Inequalities, and Number Sentences/F (.518, .645, .060)

ITBS ITEMS
with Significant Sex Differences

Mathematics Problem Solving

Level 9 (Grade 3) [25 items]

Skills Classification/ Group Favored ($\bar{p}_M$, $\bar{p}_F$, Ψ)

none

Level 12 (Grade 6) [31 items]

42. The distance between bases on a baseball diamond for older players is 90 feet. The distance between bases on a Little League diamond is $\frac{2}{3}$ as long. How many feet apart are the bases on a Little League diamond?
1) 30
2) 60
3) 135
4) (Not given)

Fraction-multiplication/M (.545, .418, .061)

46. Maple syrup is made from the sap of maple trees. If it takes 30 gallons of sap to make one gallon of maple syrup, how many gallons of maple syrup can be made from a tree that gives up 20 gallons of sap in one season?
1) $\frac{1}{2}$ 2) $\frac{2}{3}$ 3) $1\frac{1}{2}$ 4) (Not given)

Ratio and Proportion/M (.531, .400, .063)

47. The Page family drove to Montreal for their vacation. By the route they took, it was 693 miles from Tulsa to Chicago and 848 miles from Chicago to Montreal. How many miles was it from Tulsa to Montreal by this route?
1) 1541
2) 1531
3) 1441
4) (Not given)

Whole numbers-addition/F (.677, .791, .059)

52. Mike and his friend Van found a walkie-talkie set in a store for $2.59. In a catalog they found a toy telephone set they liked better for $7.95. How much more did the telephone set cost than the walkie-talkie?
1) $4.36
2) $5.44
3) $5.46
4) (Not given)

Currency-subtraction/F (.616, .745, .006)

62. The United States bought Alaska from Russia for 7.2 million dollars. The Louisiana Territory was purchased from France for 23 million dollars. How many million dollars less did the United States pay for Alaska than for the Louisiana Territory?
1) 4.9
2) 15.8
3) 30.2
4) (Not given)

Decimals-subtraction/M (.366, .268, .046)

Level 14 (Grade 8) [34 items]

73. On the outside of the garage Mr. Nelson put a basketball goal 10 feet above the driveway. The goal was $\frac{2}{3}$ the height of the garage. How many feet high was the garage?
1) $6\frac{2}{3}$
2) $13\frac{1}{3}$
3) 15
4) (Not given)

Fractions-division/M (.554, .370, .095)

75. Mr. Nelson's garage was 25.4 feet long. His car was 18.7 feet long. How many feet longer was the garage than the car?
1) 6.3
2) 7.3
3) 7.7
4) (Not given)

Decimals-subtraction/F (.685, .812, .059)

81. Paul's average score was $\frac{2}{3}$ as high as his father's. If Paul's average score was 120, what was his father's?
1) 80
2) 160
3) 180
4) (Not given)

Fractions-division/M (.454, .358, .052)

84. In the 1968 Olympic games the winner of the men's 200 meter dash was timed in 19.8 seconds. The winner of the women's 200 meter dash was timed in 22.5 seconds. The men's champion ran the 200 meter dash how many seconds faster?
1) 2.7
2) 3.8
3) 3.7
4) (Not given)

Decimals-subtraction/F (.558, .675, .054)